



Open Access for Digitization Projects (2011)

Citation

Suber, Peter. 2011. Open Access for Digitization Projects. In *Going Digital: Evolutionary and Revolutionary Aspects of Digitization*, ed. Karl Grandin, 70-93. Stockholm: Nobel Foundation.

Permanent link

<http://nrs.harvard.edu/urn-3:HUL.InstRepos:14976387>

Terms of Use

This article was downloaded from Harvard University's DASH repository, and is made available under the terms and conditions applicable to Other Posted Material, as set forth at <http://nrs.harvard.edu/urn-3:HUL.InstRepos:dash.current.terms-of-use#LAA>

Share Your Story

The Harvard community has made this article openly available.
Please share how this access benefits you. [Submit a story](#).

[Accessibility](#)

OPEN ACCESS FOR DIGITIZATION PROJECTS*

Peter Suber

When should digitization projects commit to open access (OA)?

I want to focus this question on public policy, not law or utility. If it were a question about law, the answer would be easy. As far as I know, there is no legal obligation in any country to make the results of any kind of digitization project OA. If it were a question about utility, the answer would also be easy, though the reverse. The results of a digitization project would always be more useful if they were OA.

Yet there may be good policy reasons to make some digitization projects OA even when not legally required, and there may be good reasons to change the law. Likewise, there may be good policy reasons to allow some access decisions to be made by stakeholders who will not choose OA.

Worldwide, more than 30 public funding agencies now operate on the principle that the results of publicly-funded research should be OA.¹

I started this essay to see how far I could defend the analogous principle that the results of publicly-funded digitization projects should be OA. The presence of public funding supports an OA argument in both domains. But digitization projects differ in OA-relevant respects more often than public funding agencies do, and even when they seem to be similar in all relevant respects, they frequently differ in their access policies. There is very little discernible pattern, and no matter what perspective we take, some of the policy divergence will be justified and some will not. This is a good reason to step back and think about the principles that ought to guide access policies for digitization projects.

Let me start with two relatively simple cases.

Case 1. When a digitization project uses public funds, and digitizes works in the public domain (PD), then the results should be OA.

For example, when Ontario digitizes the print editions of its historical statutes, it should provide OA to the digital editions.²

Case 2. When a project uses private funds, and digitizes works under copyright, then it should follow the wishes of the copyright holder. The results needn't be OA.

For example, when a private journal uses its own money to digitize recent back issues,

* Paper revised August 2009.

still under copyright, it needn't make them OA. It may put them online behind a paywall and sell access to them. Or it may keep them offline for its own private research purposes.

When *The Atlantic* digitized all 151 years of its backfile at its own expenses, chose to provide OA only to the most recent 15 years' worth, and toll access (TA) to the rest, then both the OA and TA parts of its project were entirely within its prerogatives.³

I'm sure you already see the supporting arguments for these two outcomes, but let me sketch them anyway. The principles behind them will help us navigate the issues in the more complicated cases.

The first case depends on the principle that public funds should be spent in the public interest. OA provides public access, and anything less than OA, or any access and usage restrictions, would compromise the public interest. The use of public funds obliges us to serve the public interest, and when we are digitizing PD works we encounter no barrier in the form of a copyright holder demanding access or usage restrictions. Taxpayers should not have to pay again for access to the digital editions. They should not pay to create an asset for the private enrichment of one citizen, one group, or one corporation, especially at the expense of the general public. Nor should they pay to create a digital asset which can only be accessed offline by the lucky few who are able to travel to a certain physical library or archive.

The second case depends on copyright law. Copyright holders have enforceable rights to their works, even if those rights are limited and temporary. Whatever the limits happen to be at a particular place and time, copyright holders should be free to exercise their rights up to the edge of those limits. They may waive or transfer their rights, of course, and it will be important that they might be asked to do so in order to enter a certain contract or use someone else's funds, especially to use public funds. But when copyright holders are using their own funds or the funds of a willing partner to digitize their own works, they should be free to offer the digital editions on any terms they please. The copyrighted backfiles of a journal might be more useful if they were OA. But I do not want to defend the idea that everything useful should be free, which would entail the abolition of copyright.

The principle of the first case leads us to applaud Ontario for providing OA to its digitized statutes, which are all in the public domain.⁴

Likewise, it leads us to criticize Oregon for falsely claiming copyright in the digital edition of its statutes and threatening to sue anyone who copied them. (This was Oregon's position until challenged by Carl Malamud in June 2008.)⁵

It leads us to criticize Pakistan for making the digital edition of its statutes freely accessible only to the country's lawyers rather than OA to all users.⁶

The British Library Digitisation Strategy 2008–2011 tells us that the BL plans to use public funds to digitize a mixed collection of PD and copyrighted works. Some of the digital editions will be OA and some will not. We can praise the library if the plan is to provide OA to the PD works. In Case 3 we will ask whether the use of public funds is enough to require OA even for works under copyright.⁷

JISC used public funds to digitize the backfiles of Oxford journals, which had already been supported by Oxford's own public funds. Whether JISC and Oxford should provide OA to issues still under copyright will be explored in Case 3. But under the principle of our first case, they should at least provide OA to any issues old enough to have passed into the PD. However, Oxford provides OA to none of the digitized backfiles – as opposed to more recent back issues which may have been OA from birth. (More below.)⁸

The principle of our second case leads us to conclude that *The Atlantic* didn't have to provide OA to any of its backfile, not even the oldest part which had passed into the PD. Its decision to provide OA to the most recent 15 years' worth is beyond the call, even if based on self-interest. Its decision to provide TA to the rest, especially to the PD issues, may prove difficult to enforce. (At least in the U.S., users may lawfully treat any copies which escape the paywall as works in the PD.) But as long as the journal avoids copyfraud, or the false claim of copyright, it should be free to try.⁹

The Dutch medical journal, *Nederlands Tijdschrift voor Geneeskunde*, is like *The Atlantic* except that it chose to provide OA to the oldest issues rather than the newest, using a five-year moving wall. Like *The Atlantic*, it paid for the digitization of its 150+ year backfile with its own funds (as far as I can tell). Like *The Atlantic*, it did not have to provide OA to any of it. Unlike *The Atlantic*, it does not have to try to restrict access to PD digital editions of PD back issues, which once online, may be copied and redistributed at will.¹⁰

Precisely because public funding pulls toward OA and private funding pulls toward the interests of the private funders, participants at the October 2007 LIBER/EBLIDA meeting in Copenhagen noted (in Recommendation 23) that "Private funding of digitisation activity may restrict access by the user, which is not in the interests of the European citizen. LIBER and EBLIDA do not see this as the preferred route for funding the digitisation of content."¹¹

Someone might object that some publicly-funded agencies should follow a cost-recovery model. The agencies have a mission to serve the public (the objection would continue), but they can best serve the public by charging for access, recovering their costs, and making their budgets go further. For example, this is the model of the Ordnance Survey, the UK mapping agency.¹²

In reply we can point out that several independent empirical studies conclude that OA stimulates significant economic activity, and that governments can generate much more revenue through taxes on that economic activity than through access fees on public data. In the case of research, this has been well-documented in several studies by John Houghton.¹³

For example, Houghton's first major study concluded that "With the United Kingdom's GERD [Gross Expenditure on Research and Development] at USD 33.7 billion and assuming social returns to R&D of 50%, a 5% increase in access and efficiency [Houghton's conservative estimate] would have been worth USD 1.7 billion; and... With the United States' GERD at USD 312.5 billion and assuming social returns to R&D of 50%, a 5% increase in access and efficiency would have been worth USD 16 billion."¹⁴

In the case of public data of the sort collected and sold back to the public by the Ordnance Survey, the UK Office of Fair Trading concluded that the cost-recovery model "cost the UK economy £500 million [per year] in lost opportunities."¹⁵

Even if Cases 1 and 2 are not themselves very simple or non-controversial, I want to use them to mark the two poles of a spectrum of cases which are even less simple. Here are three of those less simple cases.

Case 3. All the funds are public, but all the works to be digitized are under copyright.

In this case, the use of public funds pulls in favor of OA. But the copyright pulls in favor of the copyright holder. Should one side have its way at the expense of the other? If not, what compromise should we seek?

This case arises, for example, when a public agency like the US National Library of Medicine (NLM) or the UK Joint Information System Committee (JISC) funds the digitization of a journal's backfile, including issues still under copyright. When the NLM funded the digitization of the *BMJ* backfile, *BMJ* was willing to make the backfile OA without delay. The entire *BMJ* backfile to 1840 has been OA since May 2009.¹⁶

When JISC funded the digitization of the Oxford journal backfiles,

Oxford was not willing to make them OA, apparently even with a delay, although JISC did buy a license to the Oxford backfile for UK citizens. (The license will expire in July 2011. I cannot tell whether UK taxpayers, through JISC, paid once for the digitization and then paid again for the national license. Of course if the license is renewed in 2011, taxpayers will pay yet again; and if it is not renewed, they will lose their access.)¹⁷

For now let's focus on the case of a journal seeking a grant from a public funder, hoping to use the grant to digitize its copyrighted backfile and hoping to sell access to the online digital edition. It is the Oxford case, but artificially tidied up to eliminate the national license (close to OA for UK residents), TA for those outside the UK (one way in which the license falls short of OA), the limited duration of the license (another way in which it falls short of OA), and the possibility of multiple payments from the public funder.

We can imagine many kinds of compromise between the public and the rights-holding publisher. For example, we could make the works free of charge but not free for any sorts of use or reuse beyond fair use or fair dealing (see Case 5). We could make the OA copies low-res and the TA copies high-res. We could put ads on the OA copies. I mention these in order to stimulate the imagination. Over time the stakeholders may find many acceptable ways to strike the compromise, even if they also find many unacceptable ways to do it.

Here I want to focus on a compromise suggested by the analogy to publicly-funded research.¹⁸

In the case of publicly-funded research, the US National Institutes of Health (NIH) pioneered a compromise later followed by all other funding agencies with OA policies: a period of temporary exclusivity for the publisher followed by OA for the public. When NIH grantees publish articles based on NIH funding, they must deposit the peer-reviewed manuscripts in the NIH's OA repository (PubMed Central) as soon as they are accepted for publication. But the manuscripts are not made OA until after an embargo period of up to 12 months.

The delay is a compromise with the public interest, just as it is a compromise with the publisher's private interest. Because the embargo exists, publishers have a period in which to sell access to their priced editions without competition from OA editions. Because it is temporary, the public eventually gets public access to publicly-funded research.

Publishers who believe the NIH policy is not a fair compromise should seek a different compromise, for example by tweaking the embargo period, rather than demand a no-compromise position which could

deprive the public of OA for the full duration of copyright. While publishers have their reasons to lengthen the embargo, many other groups have reason to shorten it, among them researchers, practicing physicians, patients, non-profit organizations, and for-profit manufacturers. If both sides acknowledge the need for compromise, then their engagement on the length of the embargo, or on the precise terms of the compromise, is much more likely to be fruitful and constructive.¹⁹

The analogy of publicly-funded research and publicly-funded digitization should not leave the impression that the embargo compromise works the same way in both domains. We must note an often-overlooked aspect of the NIH policy. The NIH requires grantees to retain the right to authorize OA through PubMed Central. Hence, grantees are not in a position to transfer the full bundle of copyright to publishers. Publishers never acquire the right to deny permission for OA or claim infringement, and therefore cannot be called “the copyright holders” without qualification. Publishers who oppose the NIH policy understand the incompleteness of the transferred bundle of rights very well, and protest it. Nevertheless, in their lobbying rhetoric they call themselves “the copyright holders” without qualification, misleading many observers and policy-makers.²⁰

By contrast, in a digitization project we are often dealing with the full copyright holders. Nevertheless, the embargo compromise can be extended naturally to publicly-funded digitization projects.

Suppose a private journal applies to a public funder for funds to digitize its back run, and suppose that the entire back run is still under copyright. The funder would be justified in awarding the grant. At least the fact that the journal is private and under copyright need not stop it. The funder would also be justified in putting an OA condition on the grant. The grant need not require immediate OA and could allow the publisher a temporary period in which it could charge for access to the digital edition without competition from an OA edition.

More importantly, the public funder would *not* be justified in awarding the grant *without* the OA condition, or in using public funds to create a privately-owned asset which would exclude the public. Similarly, Oxford may use public funds to digitize the backfiles of Oxford journals, and it may sell access to the copyrighted issues for a temporary period. But after that the backfile must become OA.

How long should the embargo be? That should be decided by public debate and negotiation. But I have two rough criteria: First, the deal should give us OA sooner than we would otherwise have it. The public-

ly-funded digitization and OA condition will accelerate OA, while the embargo period will delay it. These conflicting trends should net out in favor of the public. If we could get OA faster some other way, then there is no reason to spend public money on the project.

Second, the longer the proposed embargo, the lower the project falls on the priority list for public funds. If the funder had to choose between two projects, one requesting a one-year embargo and another requesting a two-year embargo, then (other things being equal) it should pick the one with the shorter embargo. It might even ask the applicant proposing the shorter embargo to cut it even further to qualify for public funds. The US National Endowment for the Humanities follows the rule that, other things being equal, it will favor funding applications that promise (immediate) OA over those that do not promise OA at all.²¹

If the funder thinks a journal's proposed embargo period is too long, the journal might argue that it will still provide OA sooner than otherwise. For example, if the oldest articles it wanted to digitize would remain under copyright for another 50 years, and then it might argue that publicly-funded digitization with a 49 year embargo would give the public OA sooner than otherwise. As the copyright holder, it is in a position to insist that in the absence of public funding it will not allow OA until the expiration of copyright. The public funder need not deny the publisher's prediction or its good faith. It need only reply that it has better uses for its limited public funds than to create a 49 year monopoly for a private interest at the expense of the public.

The journal might object: "You can't require OA to our copyrighted articles!" The public funder would have several responses. "We can put conditions on our grant. You need not apply for publicly-funded grant. You can call this is an 'OA requirement' if you like, but it is really just a condition on a voluntary contract. Moreover, of course, we are a public agency and must spend our money to benefit the public."

A government would not be justified in making an *unconditional* requirement that journals provide OA to their backfiles, or at least not until it was ready to abolish copyright law. But it is fully justified in telling those who seek public funds for digitization projects, "If you take public money for this project, then you must provide OA to the results. If you don't like that, then don't take public money."

A member of the public might object: "You can't allow toll access to a publicly-funded work of digitization!" Again, the public funder would have several responses. "It is temporary. Moreover, we only funded the digitization, not the original work, and the original work is still under

copyright. But above all, in our best judgment, the public investment will make the work OA sooner than otherwise.”

Someone might object that under this rule many journals will not seek public money to digitize their copyrighted backfiles. Yes, that might happen. But it is no calamity, especially when the unpursued projects would have used public funds while excluding the public from access to the results. There is no reason why public funds should be spent on private interests unwilling to provide even delayed OA.

On the other side, for what it is worth, the prediction that many journals would rather reject both the compromise and public funds than accept both seems less likely than the opposite prediction. Allowing the private grantee a temporary period of exclusivity will invite many journals to seek public funds when an uncompromising OA principle would have scared them off.

Someone might object that I have not been consistent. I have said that copyright holders should be free to exercise their rights up to their limits (Case 2). But here I am recommending that copyright holders waive one of their rights in order to benefit from public funds.

The two positions are entirely consistent. I am not arguing that copyright holders don't have the right to insist on TA, or that they could not exercise the right if they wanted to. I am saying that they might choose to waive that right in exchange for the benefit of public funds. If they don't think it is a good deal, they don't have to take it. The deal doesn't limit their freedom; it merely offers something of value which they might or might not find worth the price of waiving their right to block delayed OA. Publishers themselves should understand this situation very well. It is exactly the kind of deal they offer to authors: give up some set of your rights in exchange for the benefit of publishing in our journal.

To obtain this kind of waiver, the public funder must deal directly with the rights-holder. The case gets more complicated when the rights-holder is not the one desiring digitization or applying for funds. For example, consider the microfiche digitization project of the publicly-funded US Education Resources Information Center (ERIC). ERIC wanted to digitize and provide OA to about 340,000 microfiche documents, some of them up to 40 years old. The documents were written by hundreds of thousands of different authors and might have hundreds of thousands of different rights-holders. Some of the documents might, after diligent inquiry, turn out to be orphans, and some might not. ERIC undertook the enormous job of trying to hunt down each copyright holder. In the end it was able to clear permissions for about 55% or 192,000 of the

documents. The rest may never be OA, despite the willingness of the U.S. Department of Education to spend public funds on their digitization.²²

Even if we adopt OA-friendly rules for orphan works, we must first go to the trouble of trying to locate the copyright holders. Otherwise we will not know whether or not the works are orphans. For more, see Case 10 in the appendix.

Case 4. The funds are provided by a public-private partnership, and all the targeted works are in the public domain.

First consider a much easier related case. If Penguin Books digitizes an early PD edition of *Pride and Prejudice* with its own funds, it should be free to sell it. It need not give it away just because the original was PD. If you agree, then it seems that public funding is a more critical variable than PD status.

The difficult case here is when we pay for the digitization of a PD work with a mix of public and private funds, a common practice. Many public funders are unable to pay for a certain project on their own, or try to stretch their budgets by recruiting private partners. The use of public funds pulls the project toward OA, and the use of private funds pulls the project toward the wishes of the private funders, which may be TA.

Consider the Digitizing American Imprints program, which is using public funds from the Library of Congress and private funds from the Sloan Foundation to digitize 100,000 PD books.²³

Another example is the Medical Journals Backfiles Digitization Project, co-sponsored by the Wellcome Library (private), JISC (public), and the US National Library of Medicine (public). The project includes some copyrighted but orphan works, which it promises to remove if the copyright holder steps forward and asks it to; more in Case 10.²⁴

A third is the World Digital Library, with public partners like 13 national libraries and UNESCO, and private partners like the Brown University Library, Yale University Library, and the Wellcome Trust Library.²⁵

The private partners in these three projects want OA as much as the public partners. That is good for the public and good for the working harmony of the partnership.

But what if the private partners oppose OA and want to sell access to the digital editions without competition from OA editions? In that case, we can use the embargo compromise that we used in the previous case. The private funder could erect a temporary toll gate on access to the digital editions.

If members of the public object that the digital editions are temporarily TA, we answer as we did in the previous case. A private funder made an essential contribution to the project and without its contribution OA would be delayed even further.

If the private funder objects that its period of exclusivity is only temporary, our replies are variations on the theme of our replies in the previous case. First, the public made an essential contribution to the project and must benefit as well. Second, the partnership is voluntary and the private partner did not have to join.

But beyond these, we have two additional replies we could not have used in the previous case. First, the private partner has no rights in these works, which we have stipulated are PD. Second, if the embargo period never expired, then for a fraction of the cost of digitization we would allow a private company to buy permanent exclusive rights to works in the PD (not the PD originals but the PD copies produced by the project).

If the private partner objects that the embargo period is not long enough to recoup its investment, and that it can't afford to take the risk of not recouping its investment, then it need not participate. If it has enough money to do the digitization by itself, without public partners, then it can proceed on its own and follow its own rules, turning this into the related, simpler *Pride and Prejudice* case. If it doesn't, then it should understand the need to allow all the investment partners to get something out of the deal.

In setting the length of the embargo, we must remember that it is a compromise with the public interest. The purpose is to give the private partners something, not everything, just as the public partners are only getting something, not everything. The compromise gives the private partners a chance to recoup their investment, not a guarantee. To give them all the time they need to recoup their investment could require a permanent embargo and eviscerate the very idea of compromise.

Someone might object that under this policy we could lose the contributions of profit-seeking private companies willing to invest in digitization projects. Yes, we could. But as before, it is no calamity to lose the chance to spend public funds on a project which excludes the public, or to lose the chance to spend public funds collaborating with those unwilling to provide even delayed OA.

Nevertheless, if governments wanted to do more to encourage the participation of private partners, without giving up on timely OA for the public, they could combine a fixed deadline on the embargo with a tax deduction for any part of the private partner's investment not recouped during the embargo period.

This could part of a larger plan to use tax deductions to get private companies to open up access to their research and data.²⁶

Public institutions taking on private partners for digitization projects should advertise their needs openly. If they accept a secret, no-bid, or unsolicited offer from a private company, they might end up agreeing to a longer embargo than necessary. Before accepting any private partners or at least any private partners who will resist OA and require a compromise, public agencies should undertake a transparent process of public consultation and competitive bidding. The rationale is simply that it is bad public policy to compromise the public interest more than necessary.

All three of the projects mentioned earlier, the Digitizing American Imprints program, Medical Journals Backfiles Digitization Project, and the World Digital Library, provide OA without any embargo at all. The private partners in all three cases came to the projects with the same purposes as the public partners, making compromise unnecessary. That is worth noting for two reasons. First, it shows that the principle here is that embargoes are permissible, not mandatory. The embargo is a compromise and is only necessary when a compromise is necessary. Second, it reminds us that the private partners in public-private partnerships don't always oppose OA.

(Conversely, public funders don't always support OA, as we have already seen in Case 1, on cost recovery, and will see again in Case 5, on the database right and sweat-of-the-brow doctrine.)

But some private partners do oppose OA. In January 2007 the U.S. National Archives and Records Administration (NARA) announced a partnership with Footnote.com. Under the deal, Footnote would digitize millions of pages of PD documents from the National Archives, including the papers of the Continental Congress and Matthew Brady's Civil War photographs. The deal gave Footnote non-exclusive rights to sell access to the digital editions for five years. During that time, the digital editions could be viewed without charge from terminals in NARA reading rooms in 16 states. After five years, the digital copies would be OA at the NARA web site.²⁷

During the five-year embargo period, Footnote's online access fees are \$1.99 per page or \$100 per year.²⁸

The non-exclusivity of the deal meant that other companies could sell access to their own digital editions, if they could make their own digital editions. But NARA is only willing to deal with Footnote. Moreover, the Footnote deal was not publicly announced until the contract was already signed and Footnote had already digitized 4.5 million PD documents.

There are several problems here. One is the length of the embargo period. Five years is very long. Footnote might argue that the combination of its money and embargo will speed up OA more than slow it down. But that seems unlikely in light of NARA's February 2006 deal with Google. Under that deal, Google funded the digitization of 101 PD films from NARA and provided immediate free online access to all of them.²⁹

In any case, the long-embargo problem is inseparable from the secret, no-bid contract problem. We will never know whether other private partners would have done the work with a shorter embargo period, lower access fees, or both.

In July 2007, NARA made an even worse deal with CustomFlix, a division of Amazon. The deal allowed CustomFlix to digitize films from the National Archives and sell DVD editions through Amazon. Members of the public who visit the NARA facility in College Park, Maryland, could copy the films without charge. In contrast to the Footnote deal, nothing in the CustomFlix contract or press release mentions an embargo period, suggesting an effectively permanent embargo.³⁰

The NARA-CustomFlix contract was secret until Rick Prelinger forced its disclosure with an FOIA request in August 2007. The contract gave Amazon perpetual non-exclusive license to sell the digital editions and gave NARA its own copies of the digital files and the right to use them in any lawful manner. Hence, it allowed NARA to provide OA at any point. But in striking contrast to the Footnote deal, NARA never promised to provide OA, on any timetable.³¹

In May 2008, NARA released a set of principles to guide its future digitization projects. Interestingly, it requires public comments on proposed private partnerships and highlights the importance of minimizing embargo periods. It seems that NARA heard the public criticism of the Footnote and CustomFlix deals and resolved to fix at least some of the problems.³²

Case 5. All the funds are public and all the works PD. So far, this is Case 1. But suppose that the host or funder wants to restrict the use of the digital editions.

Let's say that a work is gratis OA when it is digital, online, and free of charge, even if it is still subject to copyright or licensing restrictions. A work is libre OA when it is gratis OA and also free of at least some copyright and licensing restrictions. Gratis OA allows no uses beyond fair use (or fair dealing etc.), and libre OA allows at least some. Gratis OA removes price barriers and no permission barriers, while libre OA removes price barriers and at least some permission barriers.³³

Using these terms, we can restate Case 5 more succinctly: All the funds are public, all the works are PD, but the funder wants to make the digital editions gratis OA, not libre OA. The Oregon statutes from Case 1 fall under this description, and the issues raised by Oregon-type cases deserve a closer look.

Legally, the least complicated way for a digitizer to restrict use of a digital work is to keep it offline. Fair use and the public domain give us the right to use certain works in certain ways, but they don't give us the right to enter buildings where copies may be under lock and key.

But if the digitizer puts the digital edition online and still wishes to restrict usage, then the requested restrictions might have any of these four grounds:

1. *Copyright*. The work might be under copyright; but if so, we have dealt with the major issues in Case 3.
2. *Sui generis or database right*. The European Database Directive creates a kind of legal protection, outside copyright law, for databases that require substantial investment but lack the originality required for copyright. Ordinary digitization lacks the originality required for copyright, and individual examples lack the database element required for the sui generis right. But when many digitized files are brought together into a database, and when the effort requires substantial investment, then the collection may qualify for the sui generis or database right. If so, however, then Case 3 covers the major issues. If we can put an OA condition on public funds for holders of strong copyright, then we can do the same for holders the weaker sui generis right.
3. *Unenforceable request*. The online host might acknowledge that it has no legally enforceable right to restrict usage. But it might make an admittedly unenforceable request, appealing to courtesy or respect rather than law. For example, in the downloaded copies (but not the online copies) of Google-scanned PD books, Google asks users to retain attribution and avoid commercial use and automatic querying.³⁴
4. *Copyfraud*. The host might falsely claim copyright and attempt to ground its requested restrictions in copyright law.³⁵

Consider The European Library (TEL). This is an online collection of exhibits digitized from the national libraries of Europe. TEL didn't do the digitizing or set the copyright and licensing terms for the individual exhibits. It coordinates the separate efforts of the separate contributing libraries. In most cases, it doesn't even host the exhibits but links to digital editions hosted by the separate libraries.³⁶

It appears that all of the works on display through TEL were digitized with public funds, and that some of the digital editions are under copyright, some under the *sui generis* right, and some fully PD.

TEL provides no item-level rights or licensing information. See for example the image-scan of a handwritten letter from Napoleon I to Joachim Murat, King of Naples, from October 7, 1813,³⁷ or the image-scan of the Heiberg translation of the *Marseillaise* into Danish, published in Copenhagen in 1793.³⁸

TEL does provide item-level metadata, even if they don't include rights or licensing information. But the deep links to individual exhibits (which I used above) don't include the metadata. To find the metadata for the Napoleon letter or Heiberg translation, you have to locate the exhibits within this larger exhibition, click on them, and read the metadata off an unlinkable pop-up window.³⁹

But since that method doesn't tell us about rights or licensing, we can only learn the status of the Napoleon letter or Heiberg translation by consulting the TEL "terms of service", which tell us that:⁴⁰

The Conference of European National Librarians and its licensors hold the copyright for all material and all content in this site, including site layout, design, images, programs, text and other information (collectively, the "Content") held in The European Library. No material may be resold or published elsewhere without the Conference of European National Librarians written consent, unless authorised by a licence with the Conference of European National Librarians or to the extent required by the applicable law.

Even on the most charitable reading, this statement is false for many or most exhibits in the TEL. For the PD exhibits, it is entirely false. For the exhibits under the *sui generis* right, it falsely states the rights are based on copyright instead. (This matters, among other reasons, because copyright lasts more than five times longer than the *sui generis* right.) The attempted restriction on the sale and publication of the exhibits is groundless for the PD content, even if lawful for the other two categories. But TEL says that all the contents are under copyright, and none merely under the *sui generis* right and none in the PD. If it is true for some exhibits, it is copyfraud for others.

TEL might have intended the copyright statement to apply to the web site's apparatus, not to the exhibits themselves. But nothing in the statement suggests that distinction, and the clear language of the statement

(“all content in this site...”) suggests the opposite. Moreover, the absence of item-level rights and licensing information on individual exhibits forces us to turn to the general terms of service for that information. The statement might apply only to TEL-hosted content, rather than to content at the separate national libraries to which TEL merely links. But even the TEL-hosted content seems to fall into all three categories, not just the category of copyright, and in any case TEL points to the same terms of service for TEL-hosted exhibits and for library-hosted exhibits.

TEL might have intended the statement to be part of a clickwrap license, under which visitors agree to waive their rights to use and reuse any of the contents which happen to be PD. But the site does not ask users to click their assent to any licensing terms before viewing exhibits, and the terms of service claim to base the reuse restrictions on copyright, not contract. In any case, even if TEL used a clickwrap license to create a contract with the user, and even if the contract was enforceable, users who redistributed files that are actually PD would be making them available to people who were not bound by the contract.

This copyfraud creates several problems. First, for the PD content, the claimed restrictions are unenforceable. Anyone selling or publishing the digital edition of a PD work would be exercising protected rights under copyright law. Second, for content under the *sui generis* right, the copyright claim implies rights for the full term of copyright rather than the much shorter (15 year) term of the *sui generis* right.

Third and most important, the false claim of copyright might deceive or intimidate some users into giving up rights they are entitled to exercise. It inhibits the lawful and legitimate use of this valuable historical content.

Even the onerous NARA-CustomFlix contract acknowledged that “Content obtained by researchers through public access [via a NARA reading room] is in the public domain” and its uses could not be restricted.⁴¹

TEL should drop the false claim of copyright. It should acknowledge that much of its content is PD, and that users may use and reuse the PD content without restriction. If any of the exhibits are under the weaker *sui generis* right, rather than copyright, it should acknowledge that as well.

I don’t want to underestimate the difficulty of adding item-level rights information to each exhibit in a large collection. It can be one of the larger costs in a large digitization project. But if TEL can’t add accurate item-level rights information, it should at least stop using inaccurate site-level information in its place.⁴²

The TEL can’t do much more than that, since it didn’t digitize the

works in the collection. But the national libraries of Europe which participate in TEL can do more. They are using public funds to digitize PD works. Even if the EC Database Directive allows them to claim a *sui generis* right in the digital editions, they need not take advantage of the option. On the contrary, there are good policy reasons why they should not. It is hard to imagine how their purpose in trying to restrict usage could outweigh their mission to serve the public, promote access to the historical materials in their collections, and foster research, scholarship, art, education, and cultural development. (We know that the libraries' purpose is not cost-recovery, since they are already consenting to gratis OA.)

Finally, they should understand that libre OA facilitates preservation, among other forms of use and reuse. Long-term preservation requires making copies and migrating them to new media and formats to keep them readable as technology changes. Copyright and the *sui generis* right both raise the barrier to those useful copies, either by blocking them altogether or by requiring the expense or delay of seeking permission.⁴³

Until recently, Cornell University took a position roughly similar to TEL's for the PD books digitized from its library. It posted the works online, without a clickthrough license, but required users to seek permission for any commercial use. In May 2009, however, it reversed course. It acknowledged that the books are PD, stopped trying to restrict usage, and explained why in an exemplary public statement. In the statement, Cornell said it did not wish to "limit the good uses" of these works. On the contrary, it "decided it was more important to encourage the use of the public domain materials in our holdings than to impose roadblocks." Moreover, Cornell recognized that claiming the right to restrict usage was copyfraud, and that the criticism of copyfraud was justified.⁴⁴

Cornell would have been within its rights to put the digital editions behind a password, require users to assent to a clickthrough license, and then charge for access or impose usage restrictions. Likewise, it could have put the works online without a clickthrough license and made an admittedly unenforceable request to restrict usage. But in May it chose not to do either of these things, and not to rest on copyfraud either. The Cornell solution is especially commendable because Cornell is a private university. Either it used its own, private funds for the digitization or it used Google's. (Cornell has been a partner in the Google Library Project since August 2007.)

The U.S. doesn't recognize the *sui generis* database right and Cornell could not have relied on it. But even institutions in countries which do recognize the right can use the Cornell solution. They simply have to

decline to use the right available to them, and (in Cornell's words) decide to put "good uses" ahead of "roadblocks".

Cornell is a private university, but its solution is compelling even for public institutions. Indeed, if a private institution can drop copyfraud and support the full use and reuse of PD works, then public institutions using public funds should be able to do so as well.

Appendix

Here is a quick summary of the five cases I have discussed:

Case 1. All the funds are public, and all the works to be digitized are PD.

Case 2. All the funds are private, and all the works to be digitized are under copyright.

Case 3. All the funds are public, and all the works to be digitized are under copyright.

Case 4. The funds are provided by a public-private partnership, and all the works to be digitized are PD.

Case 5. All the funds are public, all the works PD, but the funder only wants to allow gratis OA, not libre OA.

Here are five more hard cases that will have to wait for another day:

Case 6. All the funds are private and all the works to be digitized are PD. So far this is the easy *Pride and Prejudice* case. But now add that the targeted works are rare, unique, or fragile.

The *Pride and Prejudice* case is easy in part because it's easy to get a copy of the print book for digitizing. If one digitization project offers the digital edition on onerous terms, then others can digitize the same book and offer their editions on more liberal terms. But the realistic odds of re-digitization plummet when the original is rare, unique, or fragile.

Consider the *Codex Leicester*, a volume of Leonardo da Vinci's handwritten journal which Bill Gates bought from Armand Hammer in 1994 for \$30.8 million. It is the only original da Vinci now in private hands.⁴⁵

Gates has been generous with its display: the original is on loan to a different museum every year; high-res photos of every page have been published in a book (a priced, printed book, not an OA book); and OA thumbnails are available online at Corbis. But as far as I know, he has not allowed OA to high-res images.⁴⁶

Is there a strong policy argument for asking a private individual like Gates to provide OA to this kind of unique PD work? If not, does the

argument become stronger if the owner is a private university like Cornell?

What if the digitization of the Dead Sea Scrolls is funded by private donors?⁴⁷

Case 7. All the funds are private and all the targeted works are PD. So far this is either the easy *Pride and Prejudice* case or the hard *Codex Leicester* case. Now add that all the targeted works will be provided to the project by a public institution, which acquired and curated them with public funds for public benefit.

A typical example is Google's project to digitize PD books from public university libraries, such as the University of Michigan library. Should Michigan put an OA condition on its collaboration with Google?

Case 8. All the funds are private, all the targeted works are PD, and all the works will be provided by an institution which has acquired and curated them at some expense. So far this is Google-Michigan case. But instead of a public institution using public funds, let it be a private institution acting for non-commercial purposes and with public subsidies through untaxed property and tax deductible contributions.

A typical example is Google's project to digitize the PD books from private university libraries, such as the Harvard and Cornell libraries. Should Harvard and Cornell put OA conditions on their collaboration with Google?

Similar issues arise when a PD digitization project is funded by private philanthropy, such as the Mellon Foundation, with no public partner.

Do the policy arguments for OA that apply to public funders also apply to all institutions with non-commercial purposes and tax breaks, even if private?

Case 9. The funds are from a public-private partnership, and the works to be digitized are PD. So far, this is Case 4. But instead of mere digitization, the project extends to editorial work and copyrighted commentary. The plan is to integrate the PD texts and the copyrighted commentary. The private partners and copyright holders want to publish the results in print books or TA web sites and oppose any attempt to make them OA, even after an embargo period.

See the NARA plan for a digital edition of papers of US Founding Fathers:

- The case for OA.⁴⁸

- The case for TA (though misleadingly called “open access”).⁴⁹
- The case for a middle position (including an embargo period followed by libre OA).⁵⁰

Case 10. Take any of the variations above in which the works to be digitized are still under copyright (for example, Cases 2 and 3). Now add the variable that they are orphan works.

Should the digitizer follow the Wellcome Library and make the digital editions OA, promising to take them down if the copyright holder steps forward and objects?⁵¹

Should it follow the Google book settlement and sell access?⁵²

For a middle position, see Peter Eckersley’s argument that all Google-digitized books, and especially the orphan works, whether based on originals from public or private institutions, should become OA after an embargo period.⁵³

If we diligently look for the copyright holders, fail to find them, and responsibly conclude that we are dealing with orphan works, then should we assume the lack of permission for OA until we have explicit consent from the copyright holders or national legislature? Or should we assume permission for OA until we have explicit dissent? Even after responsibly concluding that we are dealing with orphan works, should we adopt a compromise like an embargo period?

Or should we start to rethink the very idea of *permission* in cases like this? Normally, medical care without consent is battery, just as full-text copying of copyrighted texts without permission is infringement. But when an unconscious person is wheeled into an emergency room, and we are unable to get an explicit “yes” or “no”, then we start to talk about “implied consent” to receive care and “privilege” to render care. When diligent effort fails to turn up a copyright holder, and we are equally unable to get an explicit “yes” or “no”, then should we also start talking about implied consent and privilege? The stakes are not the same, but the consent quandary is the same. Do we only want to solve the consent quandary in matters of life and death, or might we also want solve it in matters of scholarship, research, art, culture, and education?

An earlier version of this article was published in the *SPARC Open Access Newsletter*, July 2, 2009.⁵⁴ For helpful feedback on that edition I thank Frode Bakken, Klaus Graf, Sebastian Krujatz, and Sanford Thatcher.

References

- 1 <http://www.eprints.org/openaccess/policysignup/>
- 2 <http://www.earlham.edu/~peters/fos/2009/04/oa-to-historic-ontario-legislation.html>
- 3 <http://www.earlham.edu/~peters/fos/2008/01/free-online-access-to-12-years-of.html>
- 4 <http://www.earlham.edu/~peters/fos/2009/04/oa-to-historic-ontario-legislation.html>
- 5 <http://public.resource.org/oregon.gov/>
- 6 http://www.dailytimes.com.pk/default.asp?page=2006\12\09\story_9-12-2006_pg11_4
<http://www.earlham.edu/~peters/fos/2006/12/free-access-to-pakistani-law-for.html>
- 7 <http://www.bl.uk/aboutus/stratpolprog/digi/digitisation/digistrategy/index.html>
<http://www.earlham.edu/~peters/fos/2008/08/british-library-digitization-plans-some.html>
- 8 http://www.oxfordjournals.org/access_purchase/archives.html
- 9 <http://www.earlham.edu/~peters/fos/2008/01/free-online-access-to-12-years-of.html>
- 10 <http://www.earlham.edu/~peters/fos/2007/11/dutch-medical-journal-opens-backfile.html>
- 11 Paul Ayris, "LIBER-EBLIDA Digitisation Workshop," *Liber Quarterly*, April 2008, p. 17.
<http://liber.library.uu.nl/publish/articles/000222/article.pdf>
- 12 <http://ur1.ca/5yms>
- 13 <http://www.cfes.com/EI-ASPM/>
- 14 <http://www.cfes.com/documents/wp23.pdf>
<http://www.earlham.edu/~peters/fos/2006/08/oa-increases-return-on-investment-in.html>
- 15 <http://www.freeourdata.org.uk/blog/?p=86>
- 16 <http://dx.doi.org/10.1136/bmj.b1744>
<http://www.earlham.edu/~peters/fos/2009/05/169-years-of-bmj-now-oa.html>
- 17 http://www.oxfordjournals.org/access_purchase/archives.html
- 18 I am not the first to suggest this compromise. See the High Level Expert Group on Digital Libraries, Sub-group on Public Private Partnerships, *Final Report on Public Private Partnerships for the Digitisation and Online Accessibility of Europe's Cultural Heritage*, 12010 European Digital Libraries Initiative, May 2008.

http://ec.europa.eu/information_society/activities/digital_libraries/doc/hleg/reports/ppp/ppp_final.pdf

In Recommendation #9 at p. 22, “The sub-group recommends that public domain content in the analogue world should remain in the public domain in the digital environment. If restrictions to user’s access and use are necessary in order to make the digital content available at all, these restrictions should only apply for a time-limited period.”

- 19 <http://www.earlham.edu/~peters/fos/newsletter/08-02-07.htm#embargo>
- 20 <http://www.earlham.edu/~peters/fos/newsletter/10-02-08.htm#nih>
- 21 <http://www.neh.gov/grants/guidelines/editions.html>
- 22 http://www.eric.ed.gov/ERICWebPortal/Home.portal?_nfpb=true&_pageLabel=Digitization
<http://www.earlham.edu/~peters/fos/2009/05/eric-microfiche-digitization-project.html>
- 23 <http://www.loc.gov/today/pr/2009/09-10.html>
<http://www.earlham.edu/~peters/fos/2009/01/loc-digitization-program-crosses-25000.html>
<http://www.earlham.edu/~peters/fos/2009/03/oa-book-scanning-program-from-library.html>
- 24 http://library.wellcome.ac.uk/doc_WTD037630.html
http://library.wellcome.ac.uk/doc_wtx043243.html
http://library.wellcome.ac.uk/doc_wtdv025956.html
- 25 <http://www.worlddigitallibrary.org/project/english/index.html>
<http://project.wdl.org/project/english/partners/>
- 26 <http://weblog.infoworld.com/udell/2006/06/12.html>
<http://www.earlham.edu/~peters/fos/2006/06/tax-breaks-to-support-oa.html>
<http://www.earlham.edu/~peters/fos/newsletter/07-02-06.htm#nih>
- 27 <http://archives.gov/press/press-releases/2007/nr07-41.html>
- 28 http://www.dancohen.org/blog/posts/national_archives_footnote_agreement
<http://www.earlham.edu/~peters/fos/2007/01/digitized-public-domain-docs-from-us.html>
- 29 <http://www.archives.gov/press/press-releases/2006/nr06-64.html>
<http://www.archives.gov/digitization/google-agreement.pdf>
- 30 <http://www.archives.gov/press/press-releases/2007/nr07-122.html>
- 31 <http://blackoystercatcher.blogspot.com/2007/08/national-archivesamazon-agreement.html>
http://www.panix.com/~footage/NARA_Amazon.pdf

- 32 <http://www.archives.gov/digitization/strategy.html>
- 33 <http://www.earlham.edu/~peters/fos/newsletter/08-02-08.htm#gratis-libre>
- 34 Here's an example, though the link may only work from IP addresses in the United States:
http://books.google.com.ph/books/download/Pride_and_Prejudice.pdf?id=2_S8xAws2G4C&hl=en&output=pdf&sig=ACfU3U36yL_-URJJlcMilNPjhX7LOkUgfA
- 35 http://papers.ssrn.com/sol3/papers.cfm?abstract_id=787244
- 36 <http://www.europeanlibrary.org/>
- 37 http://www.theeuropeanlibrary.org/exhibition/Napoleonic_wars/images/texts/_large/ennw22_l.jpg
- 38 <http://img.kb.dk/ma/nybrev/12-05/spalvor-mars.pdf>
- 39 http://www.theeuropeanlibrary.org/exhibition/Napoleonic_wars/
- 40 http://www.theeuropeanlibrary.org/portal/organisation/footer/termsofservice_en.html
- 41 <http://blackoystercatcher.blogspot.com/2007/08/national-archivesamazon-agreement.html>
http://www.panix.com/~footage/NARA_Amazon.pdf
- 42 Disclosure: I'm on the advisory board for TEL and have made my objections known. I'm still hoping to resolve the problem.
- 43 <http://www.ippr.org.uk/publicationsandreports/publication.asp?id=464>
<http://www.clir.org/pubs/reports/pub135/sec6.html>
<http://www.pocket-lint.com/news/news.phtml?newsId=3566>
- 44 <http://ur1.ca/6lew>
<http://www.earlham.edu/~peters/fos/2009/05/cornell-allows-unrestricted-use-of-its.html>
- 45 http://en.wikipedia.org/wiki/Codex_Leicester
- 46 http://www.amazon.com/Leonardo-Da-Vinci-Leicester-Notebook-Genius/dp/1863170812/ref=sr_1_2
[http://pro.corbis.com/Search/SearchResults.aspx?q=%20"Codex%20Leicester%20by%20Leonardo%20da%20Vinci"](http://pro.corbis.com/Search/SearchResults.aspx?q=%20)
- 47 <http://www.earlham.edu/~peters/fos/2007/11/digitizing-dead-sea-scrolls.html>
http://en.wikipedia.org/wiki/Dead_Sea_Scrolls#Digital_copies
- 48 <http://blog.librarylaw.com/librarylaw/2008/06/free-the-foundi.html>
- 49 <http://www.archives.gov/nhprc/publications/founders-report.pdf>
- 50 <http://www.earlham.edu/~peters/fos/newsletter/08-02-09.htm#update>

- 51 <http://wellcomelibrary.blogspot.com/2009/06/orphan-works.html>
- 52 <http://books.google.com/booksrightsholders/>
- 53 <http://www.eff.org/deeplinks/2009/06/should-google-have-s>
<http://www.earlham.edu/~peters/fos/2009/06/another-idea-for-building-oa-into.html>
- 54 <http://www.earlham.edu/~peters/fos/newsletter/07-02-09.htm#digitization>